

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 16, 2003, 16:55:57 ; Search time 1056.51 Seconds  
(without alignments)  
330.553 Million cell updates/sec

Title: US-09-856-070-21  
Perfect score: 60  
Sequence: 1 EELMLRLDYEE 12

Scoring table:	HL0SUM62	
	Xgapop	10.0 , Xgapext 0.5
	Ygapop	10.0 , Ygapext 0.5
	Fgapop	6.0 , Fgapext 7.0
	Delop	6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

[illegible]

Database : GenBank ; ★

1:	gb_ba:
2:	gb_ba:
3:	gb_trq:
3:	gb_trq:
4:	gb_cm:
4:	gb_cm:
5:	gb_ov:
5:	gb_ov:
6:	gb_paf:
6:	gb_paf:
7:	gb_ph:
7:	gb_ph:
8:	gb_pr:
8:	gb_pr:
9:	gb_ro:
9:	gb_ro:
10:	gb_ro:
10:	gb_ro:
11:	gb_sts:
11:	gb_sts:
12:	gb_sy:
12:	gb_sy:
13:	gb_un:
13:	gb_un:
14:	gb_vi:
14:	gb_vi:
15:	em_ba:
15:	em_ba:
16:	em_tun:
16:	em_tun:
17:	em_hum:
17:	em_hum:
18:	em_in:
18:	em_in:
19:	em_mu:
19:	em_mu:
20:	em_cm:
20:	em_cm:
21:	em_or:
21:	em_or:
22:	em_cv:
22:	em_cv:
23:	em_pat:
23:	em_pat:
24:	em_ph:
24:	em_ph:
25:	em_pl:
25:	em_pl:
26:	em_ro:
26:	em_ro:
27:	em_sts:
27:	em_sts:
28:	em_un:
28:	em_un:

```

29:  em_vi: *
30:  em_htg_hum: *
31:  em_htg_inv: *
32:  em_htg_other: *
33:  em_htg_mus: *
34:  em_htg_pln: *
35:  em_htg_rod: *
36:  em_htg_mam: *
37:  em_htg_vrt: *
38:  em_sy: *
39:  em_htgo_hum: *
40:  em_htgo_mus: *
41:  em_htgo_other: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	60	100.0	2514	4	BOVZERINA	M98498 Bos taurus
2	60	100.0	2940	9	HUMWIL2	755021 Human
3	60	100.0	3044	6	AX411074	AX411074 Sequence
4	60	100.0	3044	9	HSP7P1N	AF1521 Human mRNA
5	60	100.0	3047	6	AX440476	AX440476 Sequence
6	60	100.0	3072	9	HSM02625	AF142986 Homo sapi
7	60	100.0	3174	9	BC013603	BC013603 Homo sapi
8	60	100.0	215705	9	AL589931	AL589931 Human DNA
	57	95.0	478	9	AF189213	AF189213 Homo sapi
10	57	95.0	2701	10	MMEZR	X60671 Mus musculus
11	57	95.0	193302	2	AC087605	AC087605 Mus muscu
12	57	95.0	237792	2	AC074334	AC074334 Mus muscu
13	57	95.0	217417	2	AC022329	AC022329 Mus muscu
14	57	95.0	286508	9	AC125143	AC125143 Mus muscu
15	55	91.7	475	9	AF188897	AF188897 Homo sapi
16	53	88.3	1446	10	AF450498	AF450498 Rattus no
17	53	88.3	182230	2	AC110744	AC110744 Mus muscu
18	53	88.3	189873	2	AC125736	AC125736 Rattus no
19	50	84.3	424	9	AF139015	AF139015 Homo sapi
20	45	75.0	2545	5	AB019790	AB019790 Homo sapi
21	44	73.3	2790	1	SACNTPASE	Y10687 Sulfolobus
22	44	73.3	4875	1	SAC437617	AJ437617 Sulfolobu
23	44	73.3	113552	1	AC105419	AC105419 Homo sapi
24	44	73.3	172263	8	AP003376	AP003376 Oryza sat
25	44	73.3	186416	2	AC011226	AC011226 Homo sapi
26	44	73.3	187981	2	AC013786	AC013786 Homo sapi
27	42	70.0	182798	2	AF034831	AF034831 Oryza sat
28	42	70.0	187419	9	AF035053	AF035053 Oryza sat
29	42	70.0	237379	9	AC007946	AC007956 Homo sapi
30	41	68.3	63650	2	AC109211	AC109211 Mus muscu
31	41	68.3	771132	2	AP003820	AF004820 Oryza sat
32	41	68.3	136822	5	AC091091	AC091091 Gallus ga
33	41	68.3	143794	8	AP004305	AP004305 Oryza sat
34	41	68.3	145769	2	AC114932	AC114932 Rattus no
35	41	68.3	169047	9	AL337935	AL337935 Human DNA
36	41	68.3	173775	2	AC079933	AC079933 Trypanoso
37	41	68.3	234285	2	AF035325	AF035325 Oryza sat
38	41	68.3	268228	2	AC115703	AC115703 Mus muscu
39	40	66.7	343	8	YHSAALK2B	D12714 Candida mal
40	40	66.7	1089	1	U00498	U00498 Bacillus hi
41	40	66.7	55902	5	AF325922	AF325922 Takifugu
42	40	66.7	105596	2	AC099110	AC099110 Oryza sat
43	40	66.7	106762	2	AF034141	AF004141 Oryza sat
44	40	66.7	118389	2	AP004180	AP004180 Oryza sat
45	40	66.7	127645	2	CSJN00381	AF74636 Oryza sat

## ALIGNMENTS

## RESULT 1

HOVEZRINA	2514 bp	mRNA	linear	MAM 11-JUN-1993
DEFINITION	Bos taurus ezrin mRNA, complete cds.			
ACCESSION	M98498			
VERSION	M98498.1	GI:289407		
KEYWORDS	ezrin.			
SOURCE	Bos taurus brain cDNA to mRNA.			
ORGANISM	Bos taurus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.			
AUTHORS	1 (bases 1 to 2514)			
TITLE	Bergson, C.M., Zhao, H., Sajoh, K., Duman, R.S. and Nestler, E.J.			
JOURNAL	Ezrin and osteonectin, two proteins associated with cell shape and growth, are enriched in the locus coeruleus			
FEATURES	Mol. Cell. Neurosci. 4, 64-73 (1993)			
LOCUS	Location/Qualifiers			
CDS	1..2514			
ORIGIN	1..2514			
BASE COUNT	644 a 658 c 738 g 474 t			
ORIGIN	1..2514			
Alignment Scores:				
Prod. No.:	0.00332	Length:	2514	
Score:	60.00	Matches:	12	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
DB:	4	Gaps:	0	
US-09-856-070-21 (1-12) x B-VZ:R:NA (1 2514)				
QY	1	GlutLeuMeLeuArgLeuGlnAspTyrGluGlu 12		
DB	1184	GAGGAGCTGATGCTGAGCTTCAGGACTACGAGGAA 1219		
RESULT 2				
LOCUS	HOVWVIL2	2930 bp	mRNA	linear
DEFINITION	Human cytoovillin 2 (VIL2) mRNA, complete cds.			
ACCESSION	J05021			
VERSION	J05021.1	GI:340216		
KEYWORDS	Cytovillin; cytoovillin 2; microvillar protein.			
SOURCE	Human placenta, cDNA to mRNA.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 2930)			
TITLE	Turunen, O., Windqvist, R., Pakkanen, R., Grzeschik, K.H., Wahlstrom, T. and Vaheri, A.			
JOURNAL	Cytovillin, a microvillar Mr 75,000 protein, cDNA sequence, prokaryotic expression, and chromosomal localization			
FEATURES	J. Biol. Chem. 264 (26), 25727-25732 (1989)			
LOCUS	HOVWVIL2	3044 bp	DNA	linear
DEFINITION	Sequence 3721 from patent WO/229103.			
ACCESSION	AX411074			
VERSION	AX411074.1	GI:21443779		
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1			
AUTHORS	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.			
TITLE	Gene expression profiles in liver cancer			
JOURNAL	Patent: WO 0229103-A 3721 11-APR-2002;			
FEATURES	GENE LOGIC INC (US)			
LOCUS	AX411074	3044 bp	DNA	linear
DEFINITION	Sequence 3721 from patent WO/229103.			
ACCESSION	AX411074			
VERSION	AX411074.1	GI:21443779		
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1			
AUTHORS	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.			
TITLE	Gene expression profiles in liver cancer			
JOURNAL	Patent: WO 0229103-A 3721 11-APR-2002;			
FEATURES	GENE LOGIC INC (US)			
LOCUS	AX411074	3044 bp	DNA	linear
DEFINITION	Sequence 3721 from patent WO/229103.			
ACCESSION	AX411074			
VERSION	AX411074.1	GI:21443779		
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1			
AUTHORS	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.			
TITLE	Gene expression profiles in liver cancer			
JOURNAL	Patent: WO 0229103-A 3721 11-APR-2002;			
FEATURES	GENE LOGIC INC (US)			
LOCUS	AX411074	3044 bp	DNA	linear
DEFINITION	Sequence 3721 from patent WO/229103.			
ACCESSION	AX411074			
VERSION	AX411074.1	GI:21443779		
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1			
AUTHORS	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.			
TITLE	Gene expression profiles in liver cancer			
JOURNAL	Patent: WO 0229103-A 3721 11-APR-2002;			
FEATURES	GENE LOGIC INC (US)			
LOCUS	AX411074	3044 bp	DNA	linear
DEFINITION	Sequence 3721 from patent WO/229103.			
ACCESSION	AX4110			







Estimated insert size: 180402; sum-of-contigs estimation  
Quality coverage: 3.52 in Q20 bases; adagose-1p estimation  
Quality coverage: 3.34 in Q20 bases; sum-of-contigs estimation  
\* NOTE: this is a 'working draft' sequence; it currently  
\* consists of 26 contigs; the true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	1180:	contig	of	1180	bp	in	length	
*	1181	1280:	gap	of	unknown	length		
*	1281	2431:	contig	of	1151	bp	in	length
*	2432	2531:	gap	of	unknown	length		
*	2532	3715:	contig	of	1184	bp	in	length
*	3716	3815:	gap	of	unknown	length		
*	3816	5031:	contig	of	1216	bp	in	length
*	5032	5131:	gap	of	unknown	length		
*	5132	6686:	contig	of	1455	bp	in	length
*	6587	6686:	gap	of	unknown	length		
*	6687	7849:	contig	of	1163	bp	in	length
*	7850	7949:	gap	of	unknown	length		
*	7950	9782:	contig	of	1833	bp	in	length
*	9783	9882:	gap	of	unknown	length		
*	9883	11802:	contig	of	1920	bp	in	length
*	11803	11902:	gap	of	unknown	length		
*	11903	12924:	contig	of	1022	bp	in	length
*	12925	13024:	gap	of	unknown	length		
*	13025	15074:	contig	of	2050	bp	in	length
*	15075	15174:	gap	of	unknown	length		
*	15175	17413:	contig	of	2239	bp	in	length
*	17414	17513:	gap	of	unknown	length		
*	17514	20075:	contig	of	2562	bp	in	length
*	20076	23175:	gap	of	unknown	length		
*	20176	22035:	contig	of	1860	bp	in	length
*	22036	22135:	gap	of	unknown	length		
*	22136	24443:	contig	of	2308	bp	in	length
*	24444	24543:	gap	of	unknown	length		
*	24544	28186:	contig	of	3643	bp	in	length
*	28187	28286:	gap	of	unknown	length		
*	28287	35830:	contig	of	8604	bp	in	length
*	35891	36950:	gap	of	unknown	length		
*	36991	42771:	contig	of	5781	bp	in	length
*	42772	43871:	gap	of	unknown	length		
*	42872	56743:	contig	of	7872	bp	in	length
*	50744	50843:	gap	of	unknown	length		
*	50844	58333:	contig	of	7490	bp	in	length
*	58343	58433:	gap	of	unknown	length		
*	58434	66542:	contig	of	8109	bp	in	length
*	66543	66642:	gap	of	unknown	length		
*	66643	74480:	contig	of	7838	bp	in	length
*	74481	74580:	gap	of	unknown	length		
*	74581	85984:	contig	of	11404	bp	in	length
*	85985	86084:	gap	of	unknown	length		
*	86085	101596:	contig	of	15512	bp	in	length
*	101597	101696:	gap	of	unknown	length		
*	101697	126085:	contig	of	24489	bp	in	length
*	126086	126185:	gap	of	unknown	length		
*	126186	142027:	contig	of	15842	bp	in	length
*	142028	142127:	gap	of	unknown	length		
*	142128	183402:	contig	of	41175	bp	in	length

FEATURES	LOCATION/QUALIFIERS	SOURCE
141120	103502, 103504	183302

1. 163302  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="16"  
/clone="RP23-3512"

```

/clone= RP23-3312
/clone_lib="RPC1 mouse HAC library 23"
47411 a 42901 c 43483 q 46208 t 3299 others
BASE COUNT

```

Alignment Scores:

[illegible]

739 a	678 c	759 g	525 t
-------	-------	-------	-------

BOOKS

Alignment scores:

Prod. No.:	0.0164	Length:	2701
Score:	57.00	Matches:	11
Percent Similarity:	100.00%	Conservation:	1
Most Local Similarity:	91.67%	Mismatches:	0
Query Match:	95.00%	Indels:	0
Off:	10	Gaps:	0

US 09 856-070 21 (1-12) X MASTER (1 2701)

1 CluCluMet1ouArq1ouG1nASP1yrG1uG1n12

1128 GAGAGGTTNATGCTTGGCTGCAGACTACGACAG 1164

AC087605	18302 bp	DNA	linear	HIG 15-JAN-2001
<b>DEFINITION</b> Mus musculus chromosome 16 clone RP23-35J2, WORKING DRAFT SEQUENCE.				

**DEFINITION** Mus musculus chromosome 26 unordered pieces.

[illegible]

VERSION ACOH7605 1 01-1229982

```
KEYWORDS HTG; HTGS_PHASE1; HTGS_PHASE2
```

SOURCE: MUS MUSCULUS.

ORGANISM MUS MUSCULUS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 183302)

**AUTHORS** DOE Joint Genome Institute.

### III.F. Sequencing of Human Chromosome 16

Unpublished Request

REFERENCE 2 (bases 1 to 183302)

**DOE JOINT GENOME INSTITUTE,  
AUTHORS**

JOURNAL  
 Direct Submission  
 Submitted / 15 JAN 2001  
 Production Deadline  
 1999

CONOMO [submitted (13-JAN-2001)] PRODUCTION SEQUENCING FACILITY, DOE JOINT SUBMITTER (13-JAN-2001) MITCHELL DRIVE WALNUT CREEK CA 94598 USA

Genome Institute, 2000 Mitchell Hall  
 ---Genome Center  
 COMMENT

Center: Joint Genome Institute

Center Code: JGI

## Project Information

Project Information  
Center Project Name: 0

center clone name: RPCL-23 2512

## Summary statistics

Consensus quality: 129780 bases at least 0.40

Consensus quality: 150586 bases at least 0.30

Consensus quality: 159783 bases at least 0.20

Estimated insert size: 200300; agarose-tp estimation









